Automatic Analysis of Multimodal Requirements: A Research Preview

Elia Bruni[1], Alessio Ferrari[2], Norbert Seyff[3], and Gabriele Tolomei[2]

[1] University of Trento, CIMeC, Trento, Italy
elia.bruni@unitn.it
[2] ISTI-CNR, Pisa, Italy
{alessio.ferrari,gabriele.tolomei}@isti.cnr.it
[3] University of Zurich, RERG, Zurich, Switzerland
seyff@ifi.uzh.ch

Abstract. [**Context and motivation**] Traditionally, requirements are documented using natural language text. However, there exist several approaches that promote the use of rich media requirements descriptions. Apart from text-based descriptions these multimodal requirements can be enriched by images, audio, or even video. [**Question/Problem**] The transcription and automated analysis of multimodal information is an important open question, which has not been sufficiently addressed by the Requirement Engineering (RE) community so far. Therefore, in this research preview paper we sketch how we plan to tackle research challenges related to the field of multimodal requirements analysis. We are in particular focusing on the automation of the analysis process. [**Principal idea/results**] In our recent research we have started to gather and manually analyze multimodal requirements. Furthermore, we have worked on concepts which initially allow the analysis of multimodal information. The purpose of the planned research is to combine and extend our recent work and to come up with an approach supporting the automatic analysis of multimodal requirements. [**Contribution**] In this paper we give a preview on the planned work. We present our research goal, discuss research challenges and depict an early conceptual solution.

Keywords: Requirements analysis, multimodal requirement descriptions, similarity-based clustering, distributional semantics.

1 Introduction

Rich media requirements descriptions are used in several RE approaches to capture relevant information and to improve the needs gathering process [16,4,1,13]. These multimodal needs are often captured in early requirements elicitation steps. In later stages the captured text, audio and video information is analyzed and often transcribed into well-defined (text-based) requirements. Depending on the actual process and the project at hand, this task might be time consuming and costly.

B. Regnell and D. Damian (Eds.): REFSQ 2012, LNCS 7195, pp. 218–224, 2012.

Researchers have started to deal with the issue of multimodal information representation [11]. There are several attempts in computer vision to combine the visual and textual information in a common space. Taking inspiration from methods originally used in text processing, algorithms for search and retrieval have been built [18,5]. Enriching the images with text-based information allows a better description of images, and consequently enforces the semantic manipulation of the graphical data. Very recently, the Natural Language Processing (NLP) community has turned its attention to multimodality. However, the task is reversed: instead of using text to better describe the content of images, the images are exploited to improve word meaning [2].

The aim of our planned research is to take advantage from these results and to apply and extend these novel methods for requirements engineering. In particular, we aim to support the automatic analysis of multimodal requirements. We envision that future requirements engineering approaches support the use of various media types to describe requirements. This key information enables analysts to understand needs. However, we foresee that it is not the analyst who has to analyze these multimodal needs in the first place.

In Section 2 we discuss relevant work in the field. Section 3 presents our research goal and discusses research challenges and ideas on a conceptual solution. Finally, in Section 4 we discuss the benefits and limitations of the planned research and conclude the paper.

2 Background

Several research groups have been working on the automatic analysis of natural language requirements [10,7], e.g., by leveraging statistical approaches borrowed from the information retrieval and data mining domains. The majority of recent research aims at classifying system requirements on the basis of their pairwise similarity in order to ease their analysis [14]. Requirements are classified by domain-related topic using iterative classification algorithms, e.g., to discriminate among different categories of non-functional requirements, as proposed by Casamayor *et al.* [3]. The aim of this recent research is to partition a large set of requirements into more manageable subsets. Furthermore, contributions are concerned with the usage of typical information retrieval distance metrics to establish the similarity among two requirements. For example, Haynes *et al.* present an approach that exploits a clustering algorithm to identify common high-level customer needs expressed in natural language [8].

The ISTI-CNR and the University of Trento have a thorough experience in the discussed field. Their research on applying data mining technologies to the Web and novel algorithms to cluster information is considered to be a cornerstone of the planned research [12]. These technologies have been recently applied also to natural language requirements [6]. Furthermore, Bruni *et al.* [2] have introduced a distributional semantic model combining text- and image-based features, as a first step to enrich traditional semantic models with perceptual information. Their research has been driven by the endeavor of better satisfy psychological models discussing how we humans acquire and use semantic knowledge. This work highlights that we cannot only rely on linguistic context, but also on our rich perceptual experience.

The University of Zurich has started to intensively use multimodal requirements descriptions to support end-users in documenting needs and feedback. The iRequire approach enables end-users to document needs with the help of pictures, audio and text descriptions [16,17]. An end-user first takes a picture of a relevant environmental aspect (e.g., a picture of a bus stop). Furthermore, the end-user documents a need using text or audio recording (e.g., "*I would like to have the time shown on my mobile when the next one is coming*"). In a last step the end-user enters a rational and gives a short task description (e.g., "*I am waiting for the 25er. I would like to know if there is enough time left to buy a snack?*"). An analyst so far analyzes the gathered end-user needs manually. Early evaluations have shown that in most of the cases the gathered information allows humans to understand needs and to transcribe them into well-defined requirements [16]. However, this approach suffers from scalability issues if we consider a large number of end-user needs.

3 Automatic Analysis of Multimodal Requirements

The goal of our research is to investigate the automatic analysis of multimodal requirements. We plan to provide analysis methods and tools which support analysts in handling a large number of multimodal requirements. In the context of our research we define multimodal requirements as following: "*Multimodal requirements use different media types to represent information that needs to be combined to fully describe a particular requirement or need*". In other words, the information to fully understand a particular requirement is scattered and can be found in different sources. In our research we are not just focusing on cases where one requirement is described using only one media type. Our focus lies on requirements represented by information spread over multiple modalities instead. In the following, we identify three key research challenges (RC) and discuss how we plan to address them.

RC-1: Semantic Representation of Multimodal Requirements. Currently, multimodal requirements are manually identified and need to be understood by domain experts [16]. Our aim is to turn this activity from manual into automatic. A first cornerstone is identifying a common and integrated model for representing such multimodal needs which can be either composed of text, images, speech, video, or a combination of those. Therefore, several feature spaces might be chosen in order to capture different aspects of multimodal needs. In a first step, we intend to focus our attention on the text- and image-based channels, that only very recently have been managed to cohabit into the same feature space [2]. While tailoring the results of these novel studies to our context, we are able to define two vector models, i.e., a text-based distributional vector and an image-based distributional vector. Thereby, the idea is to represent each need as a vector with two sequential components. The textual component is a vector of fixed length representing the textual content, according to the text-based distributional model. The graphical component is another vector of fixed length representing the graphical content, according to the image-based distributional model. A particular advantage of this approach is that the text- and image-based models are independently constructed from different sources. As a first output of this research, we plan to be able to feed a model with text- and image-based needs to allow further processing.

RC-2: Similarity-Based Clustering of Multimodal Requirements. Requirements analysis includes the identification of needs that, though documented in different forms by the end-users, express similar or even the same actual need. To automate this, our strategy is based on the common-sense belief that there is a tendency for things to look more similar the more related they are [15]. We plan to explore several multimodal similarity functions that take care of multiple feature spaces (text- and image-based). Those functions will be used for discovering groups of similar needs. The actual needs could be thus extracted by analyzing the groups generated as the output of the algorithm. As the needs and resulting groups are not known *a priori*, we suggest to adopt an unsupervised technique, i.e., a clustering algorithm, to partition the needs into distinct groups. This can be done by applying a specific multimodal similarity function and would result in a group found to be related to a specific need. In addition, we foresee an environment in which we have to deal with a massive number of needs, therefore the clustering algorithm adopted to discover similar needs shall be designed to provide high efficiency for both static and dynamic load. We foresee that needs can be clustered off-line when we start collecting first needs. In this initial phase, an algorithm that is efficient on static data is desirable. In particular, we expect promising results from the Head-Tail Component (HTC) algorithm, which, in a recent work of one of the authors, has been proven to be effective for discovering groups of queries stored in Web Search Engine logs [12]. We expect to provide stable clusters of similar needs. To cope with a continuous stream of incoming new needs, these needs will be dynamically associated with the relevant cluster and new clusters will appear. This will allow us to automatically generate requirements topics (i.e., themes). Furthermore, we expect that clustering also supports requirements prioritization as the number of similar needs might indicate their importance.

RC-3: Improving Gathering and Analysis Processes. Within our research we plan to tackle both the automatic analysis of multimodal requirements itself and its consequences. We foresee that, by better understanding automatic requirements analysis with the help of first prototype approaches, we will learn more about the gathering process. For example, we might discover a general inclination towards needs composed of text and images, instead of speech. Or we might detect a correlation between the medium used for documentation and a particular group of needs (e.g., some types of need might be more naturally expressed through images, while others are easier to represent through text). All this information can be exploited to gradually improve the effectiveness of the analysis. Furthermore, this information might support us in better aligning the gathering process and analysis. We might be able to tailor the requirements gathering process and come up with new strategies to cluster together particular groups of needs. Moreover, gathering and analyzing contextual information (e.g., date, time, place) in addition to needs is another option for enhancement. Identifying correlations between gathered information might allow further process improvements. For example, if an end-user sends a need from a particular position where he already sent needs before or within a certain timeframe: this might indicate that the new need also belongs to the group of previously discussed ones. We plan a step-by-step validation of these hypotheses. This research challenge also highlights the we expect a process, which will gradually become mature in order to not only provide high quantity, but also to provide high quality requirements.

Figure 1 illustrates the envisioned conceptual solution and highlights key research issues discussed in the previous paragraphs: the gathering process (*RC-3*), the data representation issue (*RC-1*), the choices of a multimodal similarity function and similarity-based clustering algorithm (*RC-2*). Apart from work on the conceptual solution we have started a literature review. Next steps include the refinement of the conceptual solution (e.g., selection of adequate algorithms). We then plan to tailor these algorithms and to provide a tool prototype allowing the automatic analysis of needs. This prototype will be used to automatically analyze end-user needs gathered with iRequire. The evaluation results will support us in identifying issues regarding end-user needs gathering and analysis.

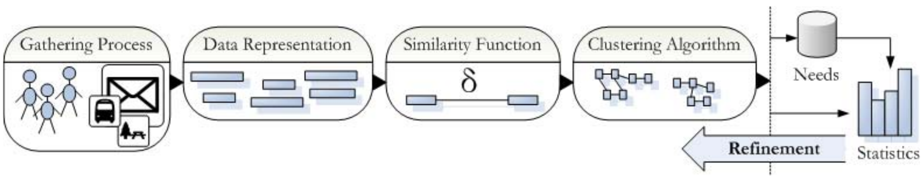

Fig. 1. Analysis of multimodal requirements: an early conceptual solution

4 Discussion and Conclusions

An important question for us is: *to what degree is automatic analysis of multimodal requirements possible?* The envisioned approach will be able to automatically group incoming needs by identifying similarity. We foresee that this will also allow establishing automated prioritization mechanisms. However, the approach will not be able to identify missing information (requirements completeness). Furthermore, we do not see the possibility to automatically detect conflicting requirements. At some point the human analyst will be needed to continue the requirements analysis. Therefore, the presentation of the automatic analysis results is a relevant issue for future work. So far we consider the discussion of research challenges and an early conceptual solution to be the first contributions of our research.

Automatic analysis of requirements might not be the only option to deal with a high number of multimodal requirements. We also consider crowd-sourcing as an option to achieve this goal [9]. However, crowd-sourcing might not ensure an independent analysis process.

Ideally, our solution will be able to analyze any kind of multimodal requirements. However, it will be necessary to tailor the method to a particular requirements gathering approach following a predefined structure for documenting multimodal requirements. We will focus on one particular gathering process and plan to support the iRequire approach [16,17]. Approaches such as iRequire can be used to gather a high number of needs requiring analysis. The discussed automated analysis mechanisms might be used within particular projects where end-users are asked to gather needs on a predefined subject (e.g., commuting). On a larger scale we also envision to analyze any end-user needs that are sent to certain receivers. With the help of automated analysis we would be able to identify needs, e.g., on novel systems, which end-user would require and which do not exist so far.

Acknowledgements. This work was partially supported by the EU-FP7-215483 (S-Cube) project, and by the PAR-FAS-2007-2013 (TRACE-IT) project.

References

1. Brill, O., Schneider, K., Knauss, E.: Videos vs. Use Cases: Can Videos Capture More Requirements under Time Pressure? In: Wieringa, R., Persson, A. (eds.) REFSQ 2010. LNCS, vol. 6182, pp. 30–44. Springer, Heidelberg (2010)
2. Bruni, E., Tran, G.B., Baroni, M.: Distributional semantics from text and images. In: Proc. of the EMNLP GEMS Workshop, Edinburgh (2011)
3. Casamayor, A., Godoy, D., Campo, M.: Identification of non-functional requirements in textual specifications: A semi-supervised learning approach. IST 52, 436–445 (2010)
4. Creighton, O., Ott, M., Bruegge, B.: Software cinema-video-based requirements engineering. In: Proc. of RE 2006, pp. 106–115. IEEE CS, Washington, DC (2006)
5. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. ACM Comput. Surv. 40(2) (2008)
6. Ferrari, A., Tolomei, G., Gnesi, S.: A clustering-based approach for discovering flaws in requirements specifications. In: Proc. of SAC 2012 (to appear, 2012)
7. Gervasi, V., Zowghi, D.: Reasoning about inconsistencies in natural language requirements. ACM TSEM 14, 277–330 (2005)
8. Hayes, J.H., Antoniol, G., Guéhéneuc, Y.G.: Prereqir: Recovering pre-requirements via cluster analysis. In: Proc. of WCRE 2008, pp. 165–174. IEEE CS, Washington, DC (2008)
9. Huberman, B.A., Romero, D.M., Wu, F.: Crowdsourcing, attention and productivity. Journal of Information Science 35(6), 758–765 (2009)
10. Kof, L., Penzenstadler, B.: From Requirements to Models: Feedback Generation as a Result of Formalization. In: Mouratidis, H., Rolland, C. (eds.) CAiSE 2011. LNCS, vol. 6741, pp. 93–107. Springer, Heidelberg (2011)
11. Louwerse, M.: Symbol interdependency in symbolic and embodied cognition. Topics in Cognitive Science 3, 273–302 (2011)

12. Lucchese, C., Orlando, S., Perego, R., Silvestri, F., Tolomei, G.: Identifying task-based sessions in search engine query logs. In: Proc. of WSDM 2011, pp. 277–286. ACM, New York (2011)
13. Maiden, N., Seyff, N., Grunbacher, P., Otojare, O., Mitteregger, K.: Making mobile requirements engineering tools usable and useful. In: Proc. of RE 2006, pp. 26–35. IEEE CS, Washington, DC (2006)
14. Palmer, J., Liang, Y.: Indexing and clustering of software requirements specifications. IDT 18, 283–299 (1992)
15. Patwardhan, S., Pedersen, T.: Using WordNet-based Context Vectors to Estimate the Semantic Relatedness of Concepts. In: Proc. of the EACL 2006 Workshop, Trento, Italy, pp. 1–8 (April 2006)
16. Seyff, N., Graf, F., Maiden, N.: Using mobile re tools to give end-users their own voice. In: Proc. of RE 2010, pp. 37–46. IEEE CS, Los Alamitos (2010)
17. Seyff, N., Ollmann, G., Bortenschlager, M.: irequire: Gathering end-user requirements for new apps. In: Proc. of RE 2011, pp. 347–348. IEEE CS, Los Alamitos (2011)
18. Sivic, J., Zisserman, A.: Video Google: A text retrieval approach to object matching in videos. In: Proc. of ICCV 2003, pp. 1470–1477. IEEE CS, Washington, DC (2003)